APPLICATION UNDER UNITED STATES PATENT LAWS

Atty. Dkt. No. PM 275327
(M#)

Invention: METHOD OF DETECTION OF ALLELIC VARIANTS OF SCA2 GENE

Inventor (s): Samir Kumar BRAHMACHARI
Shweta Chowdry Mitali MUKERJI
Sudhir JAIN

Pillsbury Madison & Sutro LLP
Intellectual Property Group
1100 New York Avenue, NW
Ninth Floor
Washington, DC 20005-3918
Attorneys
Telephone: (202) 861-3000

This is a:

- ☐ Provisional Application
- ☒ Regular Utility Application
- ☐ Continuing Application
  - ☒ The contents of the parent are incorporated by reference
- ☐ PCT National Phase Application
- ☐ Design Application
- ☐ Reissue Application
- ☐ Plant Application
- ☐ Substitute Specification
  Sub. Spec Filed ____________
  in App. No. ____/______
- ☐ Marked up Specification re
  Sub. Spec. filed ____________
  In App. No ____/______

SPECIFICATION

FIELD OF THE INVENTION

The present invention relates to a method of detection for human Spinocerebellar ataxia 2 gene variants, and more particularly their use in applications such as molecular diagnosis, prediction of an individual's disease susceptibility, and the genetic analysis of SCA2 gene in a population. The invention also provides primer and probe sequences useful in detecting these polymorphic variations in SCA2 gene and their use in diagnosis and prediction of an individual's susceptibility to SCA2 disease.

BACKGROUND AND PRIOR ART

Spinocerebellar ataxias (SCAs) are a clinically heterogeneous group of autosomal dominant neurodegenerative disorders characterized by progressive deterioration in balance and coordination. The clinical symptoms include ataxia, dysarthria, ophthalmoparesis, and variable degrees of motor weakness. The symptoms occur due to progressive neuronal loss primarily in the cerebellum but also in other parts of central nervous system. The symptoms usually begin during the third or fourth decade of life, however, juvenile onset has been identified. Typically, the disease worsens gradually, often resulting in complete disability and death 10-20 years after the onset of symptoms. Individuals with juvenile onset spinocerebellar ataxias, however, typically have more rapid progression of the phenotype than the late onset cases.

Seven disease loci have been identified to date as causing this phenotype – Spinocerebellar ataxia 1 (SCA1) (Orr et al., Nat. Genet. 4, 221-226 (1993)), SCA2 (Pulst et al., Nat. Genet. 14, 269-276 (1996); Sanpei et al., Nat. Genet. 14, 227-284 (1996); Imbert et al., Nat. Genet. 14, 285-291 (1996)), SCA3/ MJD (Kawaguchi et al., Nat. Genet. 8, 221-227 (1994)), SCA6 (Zhuchenko et al, Nat. Genet. 15,62-68 (1997)), SCA7 (David et al., Nat. Genet. 17, 65-70 (1997)), SCA8 (Koob et al., Nat. Genet. 21, 379-384 (1999)) and SCA12 (Holmes et al., Nat. Genet. 23, 391-392 (1999)). The causative mutation associated with all these disease types is abnormal expansion of trinucleotide repeat motif in their corresponding gene. The expansion of the repeat tract beyond the normal range produces premutation allele that may further expand to disease producing mutations.

The genomes of all organisms undergo spontaneous mutation in the course of their continuing evolution generating variant forms of progenitor sequences (Gusella, Ann. Rev. Biochem. 55, 831-854 (1986)). The variant form may confer an evolutionary advantage or disadvantage relative to a progenitor form or may be neutral. In some instances, a variant form confers a lethal disadvantage and is not transmitted to subsequent generations of the organism. In other instances, a variant form confers an evolutionary advantage to the species, is eventually incorporated into the DNA of many or most members of the species, and effectively becomes the progenitor form. In many instances, both progenitor and variant form(s) survive and co-exist in a species population. The coexistence of multiple forms of a sequence gives rise to polymorphisms. Several different types of polymorphisms have been reported. A restriction fragment length polymorphism (RFLP) means a variation in DNA sequence that alters the length of a restriction fragment as described in Botstein et al.,Am. J. Hum. Genet. 32, 314-331 (1980). The restriction fragment length polymorphism may create or delete a restriction site, thus changing the length of the restriction fragment. RFLPs have been widely used in human and animal genetic analyses (Donis-Keller, Cell 51, 319-337 (1987)). Other polymorphisms take the form of short tandem repeats (STRs) that include tandem di-, tri- and tetranucleotide repeated motifs. These tandem repeats are also referred to as variable number tandem repeat (VNTR) polymorphisms. VNTRs have been used in identity and paternity analysis and in a large number of genetic mapping studies.

Other polymorphisms take the form of single nucleotide variations between individuals of the same species. Such polymorphisms are far more frequent than RFLPS, STRs and VNTRs. Some single nucleotide polymorphisms (SNPs) occur in protein-coding sequences, in which case, one of the polymorphic forms may give rise to the expression of a defective or other variant protein and, potentially, a genetic disease. Examples of genes, in which polymorphisms within coding sequences give rise to genetic disease include beta.-globin (sickle cell anemia) and CFTR (cystic fibrosis). Other single nucleotide polymorphisms occur in non-coding regions. Some of these polymorphisms may also result in defective protein expression (e.g., as a result of defective splicing). Other single nucleotide polymorphisms have no phenotypic effects.

SNPs can be used in the same manner as RFLPs, and VNTRs but offer several advantages. SNPs occur with greater frequency and are spaced more uniformly throughout the genome than other forms of polymorphism. The greater frequency and uniformity of SNPs means that there is a greater probability that such a polymorphism will be found in close proximity to a genetic locus of interest than would be the case for other polymorphisms. Also, the different forms of characterized SNPs are often easier to distinguish that other types of polymorphism (e.g., by use of assays employing allele-specific hybridization probes or primers).

Spinocerebellar ataxia 2 (SCA2), which was initially described in a Cuban population (Gispert et al., Nat. Genet. 4, 295-299 (1993)), has now been reported worldwide. The human SCA2 gene has 25 exons and encompasses approximately 130kb on 12q23-24.1 region of chromosome 12 (Sahba et al., Genomics 47, 359-364 (1998)). The molecular basis of the disease is an expansion of a CAG repeat tract in exon 1 of SCA2 gene. The molecular diagnosis of clinically suspected SCA2 patients is carried out by the correct sizing of the CAG repeats at the SCA2 locus. In normal individuals this CAG repeat is not only polymorphic in length, ranging from 14 – 31 repeats with a mode of 22 repeats, but also cryptic in nature, having one or more interrupting CAA triplets. In contrast, the SCA2 disease alleles contain a pure, contiguous stretch of 34 – 59 CAG repeats.
Sanpei and Tsuji (patent CA2241173, EPO0878543 and WO 98/18920) have provided the cDNA fragments of the gene causative of spinocerebellar ataxia type 2 having a determined base sequence. Pulst and Ramos in patent WO 97/42314 have also provided the isolated nucleic acids encoding human SCA2 protein or fragments thereof and a method of diagnosis of SCA2 disease.

Tsuji and Sanpei have also patented a method for specifically diagnosing SCA2 (patents CA2232311, EPO869186 and WO 98/03679). Therein the method comprises effecting PCR by employing DNA to be tested as template and using nucleic acid primers hybridizable with the parts of the base sequences of the SCA2 gene. The diagnosis depends on the number of the CAG repeat units in the SCA2 gene, the patient with SCA2 has the number of CAG repeat units of 35 or above while the gene of a normal subject has 15 to 24 repeats, which enables the diagnosis of SCA2.

However, these methods are not useful for detecting normal individuals carrying repeats predisposed to instability and expansion (premutation alleles) as the repeat length alone would not be the correct predictor of repeat instability at SCA2 locus due to presence of varying number of CAA interruptions. The presence of interruptions within the triplet repeats has been shown to play an important role in determining stability to a number of trinucleotide repeat disorders (Chung et al., Nat. Genet. 5, 254-258 (1993); Kunst et al., Cell 77, 853-861 (1994); Eichler et al., Nat. Genet. 8, 88-94 (1994)). It has been proposed that the presence of these interruptions confers stability and their absence predisposes alleles to instability and eventual disease status.

The prior art is lacking in any method that associates the allelic variants of SCA2 gene to the disease susceptibility. The prior art is also lacking in any study that correlates the substructure of SCA2 CAG repeat with repeat instability and predisposition to the SCA2 disease. This is the first demonstration that relates to the detection of single nucleotide polymorphisms in human SCA2 gene and their use for applications such as molecular diagnosis, prediction of an individual's SCA2 disease susceptibility or otherwise, and/or the genetic analysis of SCA2 gene in a population. The novelty of present invention is in providing a method for detecting allelic variants of SCA2 gene within the human population and their association with the disease for prediction of an individual's predisposition to SCA2.

OBJECTS OF THE INVENTION

The main object of the present invention is to provide method of detection of allelic variants of human SCA2 gene.

Another object is to provide allele specific primers and probes useful for detection of allelic variants of human SCA2 gene.

Yet another object of the invention is to provide a method for establishing association of SCA2 allelic variants with disease susceptibility.

Still another object of the invention is to provide a method for screening individuals carrying SCA2 alleles predisposed to instability and expansion.

SUMMARY OF THE INVENTION

The present invention relates to allelic variants of human Spinocerebellar ataxia 2 (SCA2) gene and provides allele-specific primers and probes suitable for detecting these allelic variants for applications such as molecular diagnosis, prediction of an individual's disease susceptibility, and/or the genetic analysis of SCA2 gene in a population.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to the detection of the allelic variants of the human SCA2 gene and their utility in predicting an individual's susceptibility to the SCA2 disease.

Accordingly, the present invention provides method detection of human Spinocerebellar ataxia 2 gene variants, said method comprising the steps of :

1. designing and synthesizing oligonucleotide primers for PCR amplification of CAG repeat containing region of exon 1 of human SCA2 gene,
2. amplifying genomic DNA of SCA2 patients and normal control individuals using the above said primers,
3. sequencing the amplified PCR product and identifying sequence variations computationally by comparing it with the already existing sequence of human SCA2 gene,
4. screening normal control individuals and SCA2 patients for novel single nucleotide polymorphisms using allele specific oligonucleotide probes,
5. computing the frequencies of CC and GT haplotypes in normals and SCA2 patients,
6. establishing the association of the CC and GT haplotype with the SCA2 disease based on their frequency distribution in normals and SCA2 patients,
7. predicting the risk or susceptibility to the SCA 2 disease based on the haplotype present at the polymorphic sites in the individual tested, GT haplotype being at low risk and CC haplotype at high risk to the disease.

In an embodiment, the primers suitable for amplification of the SCA2 gene region containing one or more polymorphic sites, are selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO:5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8 and compliments thereof.

In another embodiment, the allele specific oligonucleotide probes useful for detection of SCA2 gene variants are selected from the group consisting of SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12 and the compliments thereof, wherein the polymorphic site occupies a central position of the probe.

In yet another embodiment, the length of the oligonucleotide primers and probes is in the range of 5 to 100 bases.

In still another embodiment, allelic variants of SCA2 gene have GT and CC haplotypes, Further, the invention provides a diagnostic kit for the detection of SNP haplotypes (CC/GT) comprising suitable primers and probes selected from polynucleotide sequences under SEQ ID NO: 1 to 12.

In another embodiment of the invention a nucleic acid vector may contain the allelic variants of SCA2 gene.

In an embodiment of the invention primers suitable for amplification of SCA2 gene region containing one or more polymorphic sites are provided, said primers selected from the group comprising :

a) CTC CGC CTC AGA CTG TTT TGG TAG 3' (as listed in SEQ ID NO: 1); and

b) GTG GCC GAG GAC GAG GAG AC 3' (as listed in SEQ ID NO: 2) and compliments thereof.

In yet another embodiment of the invention allele specific primers suitable for detection of allelic variants of SCA2 gene are provided, selected from the group comprising:

a) 5' CTC GGC GGG CCT CCC CGC CCC TTC GTC GTC C 3' (as listed in SEQ ID NO: 3);
b) 5' CTC GGC GGG CCT CCC CGC CCC TTC GTC GTC G 3' (as listed in SEQ ID NO: 4);
c) 5' CCT CCC CGC CCC TTC GTC GTC 3' (as listed in SEQ ID NO: 5);
d) 5' CGC CAA CCC GCG CCT CCC CGC TCG GCG CCC GC 3' (as listed in SEQ ID NO: 6);

e) 5' CGC CAA CCC GCG CCT CCC CGC TCG GCG CCC GT 3' (as listed in SEQ ID NO: 7); and

f) 5' GCG CCT CCC CGC TCG GCG CCC G 3' (as listed in SEQ ID NO: 8) and compliments thereof.

In still another embodiment of the invention allele specific probes useful for detection of SCA2 gene variants wherein the polymorphic site occupies a central position of the probe are provided, said allele specific probes selected from the group comprising:

a) 5' CCC CTT CGT CGT CCT CCT TCT CCC CCT 3' (as listed in SEQ ID NO: 9);

b) 5' CCC CTT CGT CGT CGT CCT TCT CCC CCT 3' (as listed in SEQ ID NO: 10);

c) 5' CGC TCG GCG CCC GCG CGT CCC CGC CGC 3' (as listed in SEQ ID NO: 11); and

d) 5' CGC TCG GCG CCC GTG CGT CCC CGC CGC 3' (as listed in SEQ ID NO: 12) are compliments thereof.

The allelic variants of human SCA2 gene may comprise one or more of the following single nucleotide polymorphisms as compared with the human SCA2 complete cDNA sequence in the data base (GenBank accession number U70323).

Table 1

	Site of change	**Base change**	**Amino-acid alteration**
(A)	481	G - C	Val - Leu
(B)	552	T - C	Arg - Arg

The sites of change is in accordance with the human SCA2 complete cDNA sequence in the database (GenBank accession number U70323).

The invention also provides a method of analysing a nucleic acid from an individual for the presence of base at any one of the polymorphic sites shown in Table 1. This type of analysis can be performed on a plurality of individuals who are tested either for the presence or for the predisposition to the SCA2 disease. The susceptibility to the disease can then be established based depending on the base or set of bases present at the polymorphic sites in the individuals tested.

The invention also provides oligonucleotide sequences (as listed in SEQ ID NO: 1 to 12), suitable for use as allele specific primers and probes for the detection of polymorphic sites listed in Table 1.

Further, a diagnostic kit comprising one or more of the allele specific primers or probes along with the required buffers and accessories suitable for identification of SCA2 allelic variants to establish an individual's susceptibility to SCA2 disease is also included in the invention.

Eucaryotic expressing vectors comprising a DNA sequence coding for a protein or a peptide according to the invention are new materials and are also included in the invention. Host cells, for example, cloned human cell lines, can be transformed using the new expression vectors and are also included in the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

The manner in which the above-mentioned features, advantages and objects of the invention, as well as others which will become clear, are attained and can be understood in detail, by the particular description of the invention are illustrated in the appended drawings. These drawings form a part of the specification. It is to be noted, however, that the appended drawings illustrate preferred embodiments of the invention and thereof not to be considered limiting in their scope.

In the drawing(s) accompanying this specification:

Figure 1 is a schematic representation of the two novel single nucleotide polymorphisms in SCA2 gene. The top line depicts the position of the 25 exons of the SCA2 gene. The second line shows the relative locations of the two polymorphic sites

and the CAG repeat tract in exon 1 of SCA2 gene. Both the polymorphisms are also shown in sequence context below the gene.

Figure 2 shows the distribution of CAA triplets in SCA2 CAG repeat tract of 215 normal chromosomes. Open circles represent CAG triplets and dark circles represent CAA triplets. Alleles are grouped by GT or CC haploypes and are arranged in the ascending order of the repeat length.

Figure 3 shows the frequency distribution of CAA interruptions in normal SCA2 chromosomes with GT (open bar) and CC haplotype (filled bars). Frequencies on Y-axis are the percentage of 152 alleles with GT haplotype or 63 alleles with CC haplotype.

Figure 4 shows the details of the SNP with reference ID 695871 submitted by the applicants in the SNP database (http://www.ncbi.nlm.nih.gov/SNP/),

Figure 5 shows the details of the SNP with reference ID 695872 submitted by the applicants in the SNP database (http://www.ncbi.nlm.nih.gov/SNP/).

Figure 6 shows the complete CDNA sequence of the human SCA2 MRNA submitted by pulst, S-M in the Genbank database (http://www.ncbi.nlm.nih.gov/SNP/).

Other and further aspects, features, and advantages of the present invention will be apparent from the following description of the preferred embodiments of the invention given for the purpose of disclosure. Alternative embodiments of the invention can be envisaged by those skilled in the art. All such alternative embodiments are intended to lie within the scope of this invention.

I. Novel Polymorphisms of the Invention

As a first step to the present invention, the applicants carried out the PCR amplification of CAG repeat containing region of exon 1 of the human SCA2 gene using new oligonucleotide primers. These primers were designed in accordance with the human SCA2 complete cDNA sequence submitted by Pulst, S. –M. in the data base (GenBank

accession number U70323). The sequencing of the purified PCR product revealed two novel single nucleotide polymorphisms (SNPs) in exon 1 of human SCA2 gene. It was apparent, therefore, that there is a hitherto unrecognized allele or subtype of the human SCA2 gene.

The present invention provides a sequence for the allelic variants of human spinocerebellar ataxia 2 (SCA2) gene comprising one or more of the following single nucleotide polymorphisms compared with the human SCA2 complete cDNA sequence in the data base (GenBank accession number U70323).

Table 1

	Site of change	Base change	Amino-acid alteration
(A)	481	G - C	Val - Leu
(B)	552	T - C	Arg - Arg

The sites of changes are in accordance with the human SCA2 complete cDNA sequence in the database (GenBank accession number U70323).

(The applicants have already submitted these two SNPs in the SNP database (http://www.ncbi.nlm.nih.gov/SNP/) on August 2, 2000. The first SNP at position 481 and having either a G or a C base have a reference SNP ID 695871. The reference SNP ID for the second SNP at position 552 and with T or a C base is 695872).

The first polymorphic site (A), as shown in figure 1, had either a G or a C base and is 177bp upstream of the polymorphic SCA2 CAG repeat stretch. The second polymorphic site (B) is situated 106bp upstream of the CAG repeat tract and contains either a T or a C base. While the first substitution changes the amino acid sequence from valine to leucine, the second substitution is neutral.

For example, the nucleotide sequence of the allelic variant of human SCA2 gene having polymorphic sites as listed in Table 1 may be-

5' C TCC GCC TCA GAC TGT TTT GGT AGC AAC GGC AAC GGC GGC GGC GCG TTT CGG CCC GGC TCC CGG CGG CTC CTT GGT CTC GGC GGG CCT CCC CGC CCC TTC GTC GTC CTC CTT CTC CCC CTC GCC AGC CCG GGC GCC

CCT CCG GCC GCG CCA ACC CGC GCC TCC CCG CTC GGC GCC CGC GCG TCC CCG CCG CGT TCC GGC GTC TCC TTG GCG CGC CCG GCT CCC GGC TGT CCC CGC CCG GCG TGC GAG CCG GTG TAT GGG CCC CTC ACC ATG TCG CTG AAG CCC CAG CAG CAG CAG CAG CAG CAG CAG CAA CAG CAG CAG CAG CAA CAG CAG CAG CAG CAG CAG CAG CAG CCG CCG CCC GCG GCT GCC AAT GTC CGC AAG CCC GGC GGC AGC GGC CTT CTA GCG TCG CCC GCC GCC GCG CCT TCG CCG TCC TCG TCC TCG GTC TCC TCG TCC TCG GCC AC 3'

In the above sequence the SNPs (A) and (B) are at nucleotide position 107 and 178 respectively and are shown in bold.

II. Association Analysis with the Disease

Analysis of these two SNPs in 215 normal and 50 expanded SCA2 chromosomes revealed that although four haplotypes are possible with two biallelic polymorphic systems, only two were observed, GT or CC haplotype. No GC or CT allele was detected in our sample set suggesting that either these alleles are very rare or G, T and C, C are exclusively linked to each other. The frequency of each SNP in normal and expanded SCA2 chromosomes is summarized in Table 2.

Table 2

CAG repeat size	No. of chromosomes studied (n)	Percentage GT haplotype (n)	Percentage CC haplotype (n)
Normal (18 – 31 repeats)	215	70.7% (152)	29.3% (63)
Expanded (> 32 repeats)	50	0.0% (0)	100% (50)

In 215 normal chromosomes tested, the GT and the CC haplotype was represented in 70.7% and 29.3% respectively. Further studies on expanded chromosomes revealed a highly significant (χ^2 =.76.589, p< 0.0000) difference in the distribution of the two SNPs between the normal and the expanded SCA2 chromosomes (Table 1). All the SCA2 chromosomes (n = 50) segregated with CC allele, showing that the disease chromosomes

are in complete association with the CC haplotype. In order to establish the molecular basis for the susceptibility of CC alleles for SCA2 expansion mutation, we performed the CAA interspersion analysis of SCA2 CAG repeat stretch for chromosomes with GT and CC haplotype.

Among the total of 215 control chromosomes analysed for CAA interspersion pattern, 1.8% (4/ 215) contained none, 20.9% (53/ 215) had one, 76.7% (157/ 215) had two and 0.5% (1/ 215) had three CAA interruptions (Figure 2). A marked split was observed in the number and the pattern of CAA interruptions in the alleles with GT and CC haplotype. 98% (149/ 152) of the chromosomes with GT alleles had two or more CAA interruptions while 86% (54/ 63) of the CC alleles had either one or were devoid of interruption. This difference in the number of interruptions present on GT and CC alleles as shown graphically in figure 3, is quite significant. The first 5' CAA interruption was observed at the triplet position 9 and the second at position 14 in 97.4% (148/ 152) of the GT alleles. In contrast, 73% (46/ 63) of the CC alleles had their first 5' interruption at position 14 suggesting that absence of the most proximal 5' CAA interruption. Again a significant difference in the position of the first CAA interruption was observed between the two SNP haplotypes.

When similar length normal chromosomes with GT and the CC haplotypes were compared by CAA interspersion pattern, the CC alleles were found to have less number of interruptions than the GT alleles. And this has resulted in a concomitant increase in pure CAG repeat length in chromosomes with CC haplotype. Similarly for 215 randomly selected normal chromosomes (Figure 2), the average length of the longest uninterrupted CAG repeat tract was significantly larger (one tailed t test, $p = 0.0000$) in CC alleles (13.3 repeats) as compared to GT alleles (8.03 repeats).

It has been proposed that a minimal length of pure repeats is required to initiate instability at a repeat locus. The presence of interruptions breaks the repeats into smaller repeat tracts and thus protects the repeat from instability by reducing the length of continuous uninterrupted repeats. There are evidences in case of SCA1 and fragile X syndrome that larger uninterrupted repeats are more likely to expand than cryptic repeats. This is also true for dinucleotide repeats where the degree of polymorphism for a repeat locus is

generally proportional to the length of the perfect repeat. Since 98% of the normal chromosomes with GT haplotype have two or more CAA interruptions while majority of the alleles with a single or no CAA interruptions are found to be associated with CC haplotype (Figure 3), suggests that absence of CAA interruptions between the CAG repeat tract is one of the factors contributing to repeat instability and facilitating repeat expansion in chromosomes with CC haplotype. This is further supported by the observation that the average length of the longest uninterrupted repeat tract is much longer in CC alleles (13.3 repeats) compared to GT alleles (8.03 repeats). The length of repeat variability also reduced with an increase in over all number of interruptions. For example, the length of the uninterrupted CAG repeat tract in alleles with one interruption and CC haplotype extends from 5 – 22, whereas for alleles with two or more CAA interruption and the GT haplotype, the range is 8 – 13 pure CAG repeats.

Therefore, haplotype analysis carried out using two novel SNPs suggested that both the CAG repeat length and its substructure are important parameters in the assessment of stability of SCA2 repeat alleles. The presence of CAA interruptions at SCA2 locus play an important role in determining stability to CAG repeats and their absences predisposes alleles to expansion and eventually to disease status. A complete association of CC haplotype with SCA2 expanded chromosomes and the presence of only one or no interrupting CAA triplet in control chromosomes with CC haplotype indicates that this novel allelic variant of SCA2 allele is predisposed to expansion. In other words, the absence of GT haplotype in expanded chromosomes suggests that the GT alleles are at nearly zero risk for SCA2 disease. Therefore, these SNP haplotypes in the human SCA2 gene could be used as a method of establishing individual risk to SCA2. Moreover, the presence of these two novel SNPs in very close proximity to the SCA2 repeat region also makes them very useful genetic markers in studying the origin and the evolution of SCA2 expansion mutation. The association of the CC/ GT haplotypes with the SCA2 disease was studies in an Indian population. However similar association, i.e., GT haplotype being at low risk and CC being at high risk for SCA2 disease, can be expected to hold true for other human populations also.

III. Diagnostic Kits

The invention further provides diagnostic kit comprising at least one or more allele-specific oligonucleotide as described in SEQ ID 1 to 12. Often, the kits contain one or more pairs of allele-specific oligonucleotides hybridizing to different forms of a polymorphism. In some kits, the allele-specific oligonucleotides are provided immobilized to a substrate. For example, the same substrate can comprise allele-specific oligonucleotide probes for detecting at least one or all of the polymorphisms shown in Table 1. Optional additional components of the kit include, for example, restriction enzymes, reverse-transcriptase or polymerase, the substrate nucleoside triphosphates, means used to label (for example, an avidinenzyme conjugate and enzyme substrate and chromogen if the label is biotin), and the appropriate buffers for reverse transcription, PCR, or hybridization reactions. Usually, the kit also contains instructions for carrying out the methods.

IV. Nucleic acid Vectors

Variant genes can be expressed in an expression vector in which a variant gene is operably linked to a native or other promoter. Usually, the promoter is a eukaryotic promoter for expression in a mammalian cell. The transcription regulation sequences typically include a heterologous promoter and optionally an enhancer, which is recognized by the host. The selection of an appropriate promoter, for example trp, lac, phage promoters, glycolytic enzyme promoters and tRNA promoters, depends on the host selected. Commercially available expression vectors can also be used. Suitable host cells include bacteria such as E. coli, yeast, filamentous fungi, insect cells, mammalian cells, typically immortalized, e.g., mouse, CHO, human and monkey cell lines and derivatives thereof. Preferred host cells are able to process the variant gene product to produce an appropriate mature polypeptide.

The invention further provides transgenic non-human animals capable of expressing an exogenous variant gene and/or having one or both alleles of an endogenous variant gene inactivated. Expression of an exogenous variant gene is usually achieved by operably linking the gene to a promoter and optionally an enhancer, and microinjecting the construct into a zygote. Inactivation of endogenous variant genes can be achieved by forming a transgene in which a cloned variant gene is inactivated by insertion of a

positive selection marker. The transgene is then introduced into an embryonic stem cell, where it undergoes homologous recombination with an endogenous variant gene. Mice and other rodents are preferred animals. Such animals provide useful drug screening systems.

The invention is illustrated by the following diagrams wherein :

The following examples are given by way of illustration of the present invention and should construed to limit the scope of the present invention.

EXAMPLE 1

Identification of allelic variants of SCA2 gene:

This example describes the identification of allelic variants of human Spinocerebellar ataxia 2 gene by PCR and sequencing using certain oligonucleotide primers according to the invention. DNA was extracted from human peripheral blood leukocytes using a modification of the salting out procedure. The concentration of the DNA was determined by measuring the optical density of the sample, at a wavelength of 260 nm. The DNA was then amplified by polymerase chain reaction by using the oligonucleotide primers:

1. 5' CTC CGC CTC AGA CTG TTT TGG TAG 3' (as listed in SEQ ID NO: 1) and
2. 5' GTG GCC GAG GAC GAG GAG AC 3' (as listed in SEQ ID NO: 2).

The samples were denatured at 94°C for 3 min followed by 35 cycles of denaturartion 94°C, 45sec), annealing (52°C, 30sec), extension (72°C, 45 sec) and a final extension of 7 min at 72°C in a Perkin Elmer GeneAmp PCR System 9600. This reaction produced a DNA fragment of 459bp when analysed by genescan analysis using ABI prism 377 automated DNA sequencer (459bp product had 22 repeats at polymorphic CAG repeat region). The PCR product was purified from band cut out of agarose gel using a QIAquick gel extraction kit (Qiagen) and both the strands of the PCR product were directly sequenced using dye terminator chemistry on an ABI Prism 377 automated DNA sequencer with the PCR primers. The PCR product was shown to be identical to the human ataxin-2 (SCA2) mRNA, complete cds sequence in the data base (accession number U70323), submitted by Pulst, S. –M., except for the previously mentioned two single base changes as listed in table 1.

EXAMPLE 2

Nucleotide sequence of the Allelic Variant of SCA2 gene:

The nucleotide sequence of the allelic variant of SCA2 gene derived using the method as described in example 1–

5' C TCC GCC TCA GAC TGT TTT GGT AGC AAC GGC AAC GGC GGC GGC GCG TTT CGG CCC GGC TCC CGG CGG CTC CTT GGT CTC GGC GGG CCT CCC CGC CCC TTC GTC GTC CTC CTT CTC CCC CTC GCC AGC CCG GGC GCC CCT CCG GCC GCG CCA ACC CGC GCC TCC CCG CTC GGC GCC CGC GCG TCC CCG CCG CGT TCC GGC GTC TCC TTG GCG CGC CCG GCT CCC GGC TGT CCC CGC CCG GCG TGC GAG CCG GTG TAT GGG CCC CTC ACC ATG TCG CTG AAG CCC CAG CAG CAG CAG CAG CAG CAG CAG CAA CAG CAG CAG CAG CAA CAG CAG CAG CAG CAG CAG CAG CAG CCG CCG CCC GCG GCT GCC AAT GTC CGC AAG CCC GGC GGC AGC GGC CTT CTA GCG TCG CCC GCC GCC GCG CCT TCG CCG TCC TCG TCC TCG GTC TCC TCG TCC TCG GCC AC 3'

In the above sequence the two SNPs as given in Table 1 are at nucleotide position 107 and 178 respectively and are shown in bold.

EXAMPLE 3

GT alleles are at nearly zero risk for SCA2 diseases:

A method as described in example 1 is applied to a series of DNA samples extracted from Spinocerebellar ataxia 2 positive individuals and normal controls. There is observed a statistically significant difference ($p < 0.0000$) in the frequency distributions of the SNP haplotypes generated using the single nucleotide polymorphisms in normal and expanded SCA2 chromosome. The results obtained are summarized in the table below:

Diagnosis	SCA2 haplotype	
	GT	CC
Control Individuals	70.7%	29.3%
Spinocerebellar ataxia 2 Patients	0.0%	100.0%

A complete association of CC haplotype with SCA2 disease chromosomes indicates that SCA2 alleles with the CC haplotype are predisposed to expansion. In other words, the absence of GT haplotype in expanded chromosomes indicates that GT alleles are at nearly zero risk for SCA2 disease. Therefore, these SNP haplotypes in the human Spinocerebellar ataxia 2 gene could be used as a method of establishing individual risk to Spinocerebellar ataxia 2. The association of the CC/GT haplotypes with the SCA2 disease was studies in an Indian population. However similar association, i.e., GT haplotype being at low risk and CC being at high risk for SCA2 disease, can be expected to hold true for other human populations also.

EXAMPLE 4

Allele specific primers used for the detection of the allelic variants of SCA2 gene:

1. 5' CTC GGC GGG CCT CCC CGC CCC TTC GTC GTC C 3' (as listed in SEQ ID NO: 3)
2. 5' CTC GGC GGG CCT CCC CGC CCC TTC GTC GTC G 3' (as listed in SEQ ID NO: 4)
3. 5' CCT CCC CGC CCC TTC GTC GTC 3' (as listed in SEQ ID NO: 5)
4. 5' CGC CAA CCC GCG CCT CCC CGC TCG GCG CCC GC 3' (as listed in SEQ ID NO: 6)
5. 5' CGC CAA CCC GCG CCT CCC CGC TCG GCG CCC GT 3' (as listed in SEQ ID NO: 7)
6. 5' GCG CCT CCC CGC TCG GCG CCC G 3' (as listed in SEQ ID NO: 8)

EXAMPLE 5:

Allele specific oligonucleotide probes used for detection of the SCA2 gene variants:

1. 5' CCC CTT CGT CGT CCT CCT TCT CCC CCT 3' (as listed in SEQ ID NO: 9)
2. 5' CCC CTT CGT CGT CGT CCT TCT CCC CCT 3' (as listed in SEQ ID NO: 10)
3. 5' CGC TCG GCG CCC GCG CGT CCC CGC CGC 3' (as listed in SEQ ID NO: 11)
4. 5' CGC TCG GCG CCC GTG CGT CCC CGC CGC 3' (as listed in SEQ ID NO: 12)

EXAMPLE 6:

Nucleic acid vectors containing the SCA2 variant sequences:

Expression vectors and host cell transformed with the allelic variant of SCA2 gene containing one or more polymorphic sites as listed in table 1, can be prepared, for example, as detailed below.

Allelic variant of SCA2 gene can be expressed in an expression vector in which the variant gene is operably linked to a native or other promoter. Usually, the promoter is a eukaryotic promoter for expression in a mammalian cell. The transcription regulation sequences typically include a heterologous promoter and optionally an enhancer, which is recognized by the host. The selection of an appropriate promoter, for example trp, lac, phage promoters, glycolytic enzyme promoters and tRNA promoters will depend on the host selected. Commercially available expression vectors can also be used.

The means of introducing the expression construct into a host cell varies will depend upon the particular construction and the target host. Suitable means include fusion, conjugation, transfection, transduction, electroporation or injection. A wide variety of host cells can be employed for expression of the variant gene, both prokaryotic and eukaryotic. Suitable host cells include bacteria such as E. coli, yeast, filamentous fungi, insect cells, mammalian cells, typically immortalized, e.g., mouse, CHO, human and monkey cell lines and derivatives thereof. Preferred host cells are able to process the variant gene product to produce an appropriate mature polypeptide.

ADVANTAGES:

The invention shall be useful to establish genotype or base variations of SCA 2 gene. The information may be useful for molecular diagnosis, prediction of an individual's disease susceptibility to SCA2, prognosis and/or the genetic analysis of SCA2 gene in a population. The frequency of these variants can also be used to predict the prevalence of SCA 2 disease among various populations.